JOURNAL OF COMBINATORIAL THEORY 7, 331–341 (1969)

Null Genus Realizability Criterion for Abstract Intersection Sequences

GEORGE K. FRANCIS

University of Michigan, Ann Arbor, Michigan 48104

Communicated by Frank Harary

Received May 15, 1968

ABSTRACT

The plane realizability problem for an abstract signed intersection sequence is solved by a twelve-instruction finite-decision algorithm.

HISTORICAL REVIEW OF THE PROBLEM

In an empirical contribution to the analysis situs of curves Gauss [2] associated an algebraic symbol with a regular, closed plane curve that has a finite number of simple, transverse self-intersections called *nodes* (*Knotenpunkte*). He made a *word* by listing the node labels in the order in which they are encountered by some parameter traversing the curve. For example, Gauss associated the word $P\,Q\,R\,P\,Q\,R$ with the trefoil in Figure 1. He classified all words composed of up to five letters. The

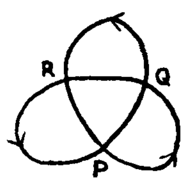

FIGURE 1

realizable words were those corresponding to some actual closed curve. He discovered that for the realizable words the two-place numbers of a letter had opposite parity.

This fact was proved by Nagy [4] to be necessary for the realizability of any word on the sphere, and so, equivalently, on the Euclidean plane.

The word $P\,Q\,P\,Q$, while not realizable in the Euclidean plane, is realizable in the projective plane. Also $P\,Q\,P\,Q$ is realizable on the torus. (Imagine a hyperbola with one branch deformed so that it intersects the other in two places.) Nagy observed that the parity criterion also sufficed for the realizability of words with fewer than five letters. The word $P\,Q\,R\,S\,T\,R\,S\,P\,Q\,T$ is not realizable in the plane even though the criterion is fulfilled (see Fig. 2). Nagy's principal device was a decompo-

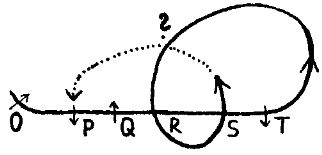

FIGURE 2

sition of the oriented one complex such a curve describes into two classes of simple, oriented cycles by painting alternating segments one of two different colors. Recently, Treybig [7] developed a more refined method of decomposing plane projections of polygonal knots into cycles which include those of Nagy as special cases. In this way he found additional necessary conditions for the realizability of words, which in the end proved also sufficient. More recently Marx gives a different necessary and sufficient condition based on the theory of planar graphs in the *Proc. Amer. Math. Soc.* (to appear; see Abstract 663-189 *Notices Amer. Math. Soc.* **16**, 1 (1969).).

A realizable word usually corresponds to several curves that are not isotopic in the plane. For example, $P\,P$ is the word for the figure-eight as well as for the double loop with index either $+2$ or -2. For this reason, Whitney [9] introduced signs at each node corresponding to the orientation of the two velocity vectors. He proposed the convention that the starting point of the parameter be on the boundary of the unbounded complementary component of the curve. He localized the orientation of the curve at the starting point by naming the orientation of velocity vector with the normal vector (pointing into the unbounded component). Thus, since Whitney, a normal immersion of the circle, or *normal loop*, is a regular, periodic, complex valued map of one real parameter, having a finite number of transverse self-intersections of multiplicity two.

Titus [5] laid the foundations for the study of the topologic properties of normal loops by means of the combinatorial properties of their *intersection sequences*. For the purpose of this paper we fix on one of several equivalent definitions. This one is a development of the definition given in Part II of Titus [6].

Definition of the Intersection Sequence of a Normal Loop

Let g be a regular C^1 map of the reals reduced modulo $2n + 2$ to the oriented Euclidean plane,

$$g : R/\text{mod}(2n + 2) \to E^2,$$

having n self-intersections occurring at integral parameter values j, so that $g^{-1}g(j)$ has one other member, denoted by j^*. The starting point, $g(0) = g(2n + 1)$, shall lie on the boundary of the closure of the unbounded complementary component of the image of g, symbolically: $g(0) \in \partial C_\infty[g]$. The orientation of $[g]$ assigns to the nodal parameter j the signum of the determinant of $dg(j^*)/dx$ with $dg(j)/dx$, denoted in this paper by

$$\nu(j) = \text{sgn}(g'(j^*) \wedge g'(j)).$$

For example, in Figure 2, where $g(3) = R$ and $g(4) = S$, $\nu(3) = +1$ and $\nu(4) = -1$.

The sign of 0 is more complicated to define. For technical reasons, which will becomes apparent later, we choose to modify the concept of a normal loop g near its starting point. Instead of smoothly closing up g at $g(0) = g(2n + 1)$, we introduce a further node at these parameter values in such a way that the modification h of g shall now be a regular arc on the interval $[-1, 2n + 2]$. The distinct end-points, $h(-1)$ and $h(2n + 2)$, shall lie in C_∞, and $h(0) = h(2n + 1)$ shall be a node. For the rest, $h = g$, say on $[\frac{1}{2}, 2n + \frac{1}{2}]$. See Figure 3. The reader may convince

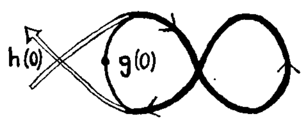

FIGURE 3

himself that the sign of the outside starting point of g, defined variously in [1, 3, 5, 8, 9], is precisely $\nu(0) = \text{sgn}(h'(2n + 1) \wedge h'(0))$. The modification of g to h is unambiguously reversible. We therefore identify the concept of an outside starting normal loop of n nodes with that of a normal arc with outside end-points of $n + 1$ nodes.

Definition of an Abstract Intersection Sequence

The triple $S = \{n, *, \nu\}$ consists in a fixed-point-free involution $*$ on the $2n + 2$ letters $0, 1, 2, 3, \ldots, 2n + 1$, such that $0^* = 2n + 1$; and in a sign ν satisfying $\nu(j^*) = -\nu(j)$. We may represent such an abstract intersection sequence by augmenting Gauss's word front and back by one new letter, and placing the sign $+$ or $-$ above each letter, so that letter pairs are oppositely signed. The trefoil in Figure 1 becomes

$$\overset{+}{O}\overset{+}{P}\overset{-}{Q}\overset{+}{R}\overset{-}{P}\overset{+}{Q}\overset{-}{R}\overset{-}{O}.$$

In this way, two normal loops with like intersection sequence differ by orientation preserving diffeomorphisms of the circle and of the plane [6, Theorem 3]. In [5] Titus found a more inclusive necessary condition for the realizability of an intersection sequence. In our terminology, this condition reads:

(*T*) For all $j < j^*$,

$$\sum \nu(a) \mid a < j < a^* < j^* = \sum \nu(b) \mid j < b < j^* < b^*.$$

The parity condition of Gauss and Nagy now reads:

(*N*) For all j, $\quad j - j^* = 1 \pmod 2$.

Of course, (*T*) implies (*N*). Not every signing of a realizable sequence satisfying (*N*) remains realizable. For example (Fig. 4) $P\,Q\,R\,S\,Q\,P\,S\,R$ is

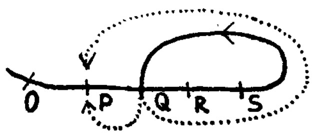

FIGURE 4

realizable if indices 1 and 2 have opposite sign, but not if they have like sign. The unrealizable sequence on five letters, already cited, can be signed to satisfy (*T*):

$$\overset{-}{O}\overset{+}{P}\overset{-}{Q}\overset{+}{R}\overset{-}{S}\overset{+}{T}\overset{-}{R}\overset{+}{S}\overset{-}{P}\overset{+}{Q}\overset{-}{T}\overset{+}{O}.$$

Nevertheless this sequence is realizable on both the torus and on the projective plane (see Fig. 5).

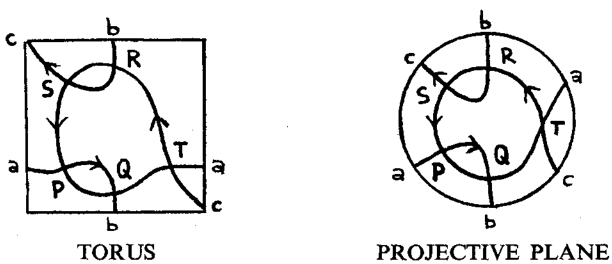

FIGURE 5

A Realizability Algorithm for Signed Intersection Sequences

In this paper we present a solution to the plane realizability problem for an abstract signed intersection sequence S in the form of a twelve-instruction finite-decision algorithm ALG(S).[1] The algorithm is given in the compact style of a flowchart (Fig. 6). This avoids the customary misery associated with multiple subscripting and notational redundancy in traditional expositions of complicated algorithms. The reader unfamiliar

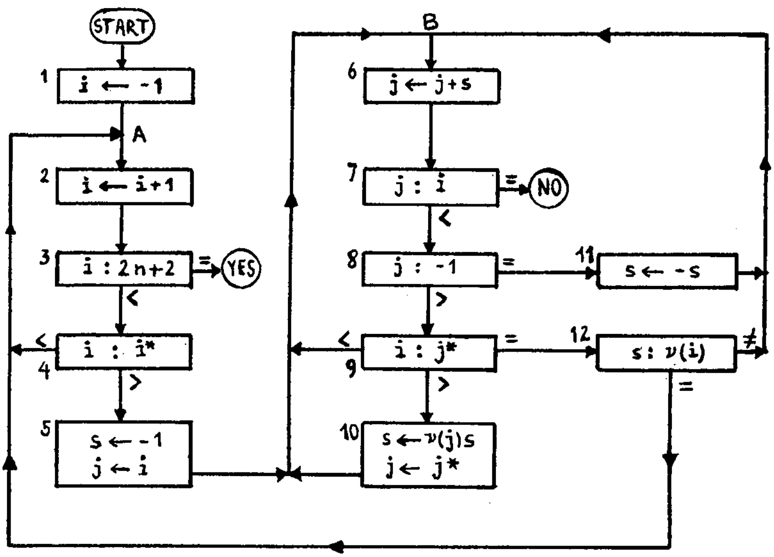

Figure 6. Flowchart for ALG(S)

Input: function $*$: $\{0, 1, \ldots, 2n+1\} \to \{0, 1, \ldots, 2n+1\}$,
ν: $\{0, 1, \ldots, 2n+1\} \to \{-1, +1\}$,
satisfying: $n \geqslant 0$, $0^* = 2n+1$, $j^* \neq j$, $j^{**} = j$, $\nu(j^*) = -\nu(j)$.

Output: YES, NO

Ranges: $i \in \{-1, 0, 1, \ldots, 2n+2\}$, $j \in \{-1, 0, 1, \ldots, 2n+1\}$, $s \in \{-1, +1\}$.

with this notation is reminded of the logical distinction between the indicative and the imperative mood in such statements as "$a = b + c$". The specification instruction "$j \leftarrow j + 1$" commands that the quantity

[1] The author is indebted to C. J. Titus for having introduced him to this subject, and for persistently drawing normal loops in precisely the way that led to the discovery of this algorithm. Thanks are due to R. F. Verhey for contributing to the simplification of the algorithm.

hitherto denoted by j be augmented by 1, and that this sum be subsequently denoted by j. The only other type of instruction used is a branching alternative. The instruction "according to the order relation that i has to j^*, do . . ." is abbreviated by "$i : j^*$."

As will be apparent in the proof below, this algorithm could be modified to display the geometric realization on an automatic plotter, provided the plotter be instructed to compensate for the inevitable "crowding" of curves that will occur due to the discrete grid used by the digital computer. Admittedly, the demonstration of the correctness of the algorithm would suffice to establish that the algorithm itself is well formed (each instruction can be executed) and finite (no infinite loops can develop), granted the consistency of geometry with number theory. However, we satisfy the demands of a "compiler" ignorant of geometry by giving an intrinsic demonstration of this in the appendix.

Proof for the Validity of the Algorithm

Consider the following typical test situation $H(a)$, indexed by $a = 0, 1, 2, \ldots, 2n + 5$.

$H(a)$: Suppose that g is a normal arc in the plane, parametrized by the interval $[-1, a + 1]$. The nodes of g occur at some of the integral values b, $0 \leqslant b \leqslant a$. We write b^* for the other value of $g^{-1}g(b)$. A sign is given to each such nodal parameter by $\nu(b) = \operatorname{sgn}(g'(b^*) \wedge g'(b))$. The starting point $g(-1)$ lies on $\partial C_\infty[g]$. Because the other end-point $g(a + 1)$ is not a node, it lies in the closure of an unambiguously defined complementary component of $[g]$, which we denote by $C_a[g]$. Suppose, further, that the collection

$$S_a = \{0, 1, 2, \ldots, a; *, \nu\}$$

is a subset of an abstract intersection sequence $S = \{n, *, \nu\}$. (The two functions operate by restriction.) Suppose, finally, that under the above circumstances, the algorithm has successfully arrived at check-point (A) in the flowchart, and is about to (re)enter instruction box (2) with the value of $i = a$.

The reader is advised to look at the flowchart. After instruction box (2), the new test variable reads $i = a + 1$. If in box (3) $i = 2n + 2$, then in situation $H(a)$ we have $a = 2n + 1$ and therefore $S_a = S$. Consequently g is a realization of S, and S is the intersection sequence of g. Otherwise $i \leqslant 2n + 1$ and there exists a value i^* in S. If in box (4) $i < i^*$, we may extend g some way into the interior of $C_a[g]$, parametrizing this extension on $[a+1, a+2]$. The combined arc on $[-1, a+2]$ satisfies $H(a+1)$. Otherwise $i^* < i$ and $g(i^*)$ is a simple point somewhere on $[g]$.

The J-subroutine, instructions (5) through (12), is designed to check whether or not it is possible to construct situation $H(a+1)$ in this case also. It would be necessary to extend g by a simple arc k, parametrized on, say, $[i-1, i+1]$, so that $k(i-1) = g(i)$, $k(i) = g(i^*)$, $\nu(i) = \operatorname{sgn}(g'(i^*) \wedge k'(i))$, and $[k] \cap [g] = \{g(i), g(i^*)\}$ only! In this way, reparametrizing the combined arc near $g(i) = k(i-1)$ only, would constitute $H(a+5)$. The subroutine achieves this by generating a sequence of pairs (s, j), $s = \pm 1$, $-1 \leqslant j \leqslant i$, which describes a search path along ∂C_a, so oriented as to keep C_a to the right of the direction traveled. Each time a value (s, j) appears at the check-point (B) one should think of the test parameter as having left the point $g(j)$ on ∂C_a, in direction $sg'(j)$, with C_a to the right of the path, about to hit $g(j+s)$ on ∂C_a with C_a to the right of $sg'(j+s)$. After box (6) we are examining the new point, $g(j)$, and the direction s, to see what must be done next in order to remain on ∂C_a with C_a to the right. For this purpose, the test path must begin in box (5) with $(-1, i)$.

If the proposed extension k were in fact possible, the test value (s, j) would read $(\nu(i), i^*)$ when it encounters k. This can be seen as follows. Since k would cross exactly at $g(i^*) = k(i)$, and $k'(i)$ points out from C_a (which is to the right of the test path), j must have value i^* and s be such that $k'(i)$ points to the left of $sg'(i^*)$, hence

$$+1 = \operatorname{sgn}(sg'(i^*) \wedge k'(i)) = s\nu(i).$$

This explains the need for $s = \nu(i)$ in box (12) for the exit to (A). Consequently, if in box (7) we find that $j = i$, we have in effect checked the entire ∂C_a and not found an exit back to (A).

It is, of course, possible for $C_a = C_\infty$. So it is possible to reach the "trunk station" $g(-1)$ out on a peninsula extending into C_∞ before finding the exit or deciding that there is none. Instructions (8) and (9) return the value $(+1, -1)$ to (B). In this way, $g(-1+1) = g(0)$ is still on ∂C_a and C_a is to the right.

Suppose we arrive at an integral point $g(j)$ on ∂C_a which is neither $g(-1)$ nor $g(i)$, with C_a to the right of $sg'(j)$. If $i < j^*$ in (9), then $g(j)$ is not a node of $[g]$ and we may continue in the same direction, remaining on ∂C_a and keeping C_a to the right. If $i > j^*$, then $g(j)$ is a node. Because C_a was to the right entering this point, we remain on ∂C_a with C_a still to the right if we make a right turn, i.e., leave $g(j^*)$ in the direction $s_1 g'(j^*)$, where

$$+1 = \operatorname{sgn}(s_1 g'(j^*) \wedge sg'(j)) = s_1 s\nu(j).$$

For this reason, box (10) returns the value $(j^*, s\nu(j))$ to (B). Finally, if $i = j^*$, we have seen above that it remains to check whether $s = \nu(i)$ in (12). If it does not, we treat this point as in the case $i < j^*$. It would be fallacious to conclude immediately that $(\nu(i), i^*)$ will not be reached. The unconvinced reader should examine situation $H(6)$ in the sequence

$$\overset{-}{O}\overset{+}{P}\overset{-}{Q}\overset{+}{R}\overset{-}{P}\overset{+}{Q}\overset{-}{S}\overset{+}{S}\overset{-}{R}\overset{+}{O}$$

as represented in Figure 7.

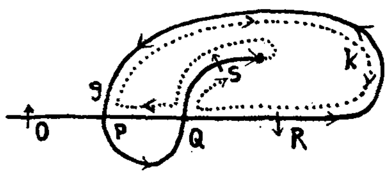

FIG. 7

In the event of the subroutine exiting at (12) it is apposite to establish the constructibility of k. Had we saved the sequence (s, j), say by collecting it at (B), we could specify the construction of k as follows. Begin at $g(i)$, make a positively oriented U-turn into the interior of C_a, and build k parallel to ∂C_a, a small distance to the right of the test path. Recall that we turned right at a node; hence k will make a right turn and thus remain parallel to ∂C_a and in the interior of C_a. An exception to this rule must be made in the event that $g(-1)$ is reached, so as to prevent k from enclosing this point in a bounded component. In this event, $C_a = C_\infty$, and we can trace a very large detour all the way around the figure drawn so far, so as to return to the other side of the peninsula and continue along that side back to the mainland. Done carefully, so as not to produce extraneous intersections, we can leave $g(-1)$ clear and outside. When we come to $g(i^*)$, turn k so as to cross g transversally and continue on a little way. Since ∂C_a is (essentially) simple, so is k. Now g and k can be reparametrized and we reach situation $H(a+1)$. Needless to say, we should have made all our corners on k smooth enough to support a regular parameter. To achieve complete mathematical rigor in this construction, we would avail ourselves of the theory of normal tubular neighborhoods and piecewise normal arcs developed and used in [1].

By finite induction on the index a in the situations $H(a)$, if the algorithm terminates at YES for a given sequence, then it has also been geometrically realized. (It is clear that the interval $[-1, +1]$ satisfies $H(0)$.) If, on the other hand, S is realizable, and is the sequence of the normal loop h, we achieve situation $H(a)$, for each a, by setting $g = h \mid [-1, a + \frac{1}{2}]$, reparametrized to run on $[-1, a + 1]$ by compressing the last quarter segment, and setting $k = h \mid [a + \frac{1}{2}, a + \frac{3}{2}]$, appropriately reparametrized. From the analysis of the algorithm above, it follows that it will not terminate at N0, at least not in this cycle from $H(a)$ to $H(a + 1)$. Inductively then, the algorithm will terminate at YES.

Proposed Questions for Future Research

Problem 1

It is clear that, by the introduction of a bridging handle whenever the need arises to "get out of a trap," every signed (or unsigned) intersection sequence is realizable on the surface of sufficiently high genus. For S a signed intersection sequence, let us define the *genus* $\gamma(S)$, to be the minimum genus on which S is realizable. We have shown that $\gamma(S) = 0$ if and only if $\mathrm{ALG}(S) = \mathrm{YES}$.

Moreover, $\mathrm{ALG}(S)$ can be incorporated into an algorithm, $\mathrm{ALG}^0(S^0)$, that decides the plane realizability of an unsigned sequence S^0. A simple, but inefficient program would be to check $\mathrm{ALG}(S)$ for each of the 2^{n+1} possible signings ν of S^0. Matters could be improved by selecting only those ν for S^0 for which condition (T) holds. It is not clear that this is much of an improvement. An immediately transparent and elegant method is as follows.[2]

Suppose we are testing (S^0, μ) for realizability in ALG. We are testing the i-th stage. If we pass to NO, never having gone through box (12) at any time in that subroutine, then (S^0, ν) is not realizable for any ν with $\nu(j) = \mu(j), j \leqslant i$. On the other hand, if we come to NO having had to pass through (12), change the signing ν to λ, by setting $\lambda(i) = -\mu(i)$, but leaving $\lambda(j) = \mu(j)$, $i \neq j \neq i^*$. It is clear that (S^0, λ) would have weathered the test at the i-th stage successfully. We return the data to (A) and continue using λ.

We may now ask for an algorithm that computes $\gamma(S)$ generally. For an unsigned sequence S^0, we define the genus $\mid \gamma \mid$ to be the minimum of $\gamma(S)$ over all possible signings S of S^0.

---

[2] The author is indebted to the referee for suggesting this, and wishes to express his appreciation to the referee for his several helpful observations and improvements of the rest of this article.

Problem 2

We can define the genus of the various necessary conditions for realizability as arithmetic functions of the number of nodes n. Let

$$\Gamma_N(n) = \max \mid \gamma \mid (S^0)$$

over all unsigned sequences S^0 of n nodes that also satisfy the parity condition (N). As we have seen, Gauss found that $\Gamma_N(n) = 0$ for $n = 1, 2, 3, 4$ and that $\Gamma_N(5) \geqslant 1$. Is $\Gamma_N(5) = 1$? Define $\Gamma_T(n) = \max \gamma(S)$ over all signed sequences S of n nodes that also satisfy Titus's condition (T). Since (T) implies (N), $\Gamma_T(n) \geqslant \Gamma_N(n)$. It is an equality only for $n = 1, 2, 3$ and fails already for $n = 4$. Since both functions are monotonic, one might ask for their difference.

Problem 3

It is true that if a signed sequence is realizable on the torus it is also realizable on the projective plane and vice versa? See Figure 5. What, if any, relation do the topologic invariants of a compact two surface have to the algebraic structure of the sequences realizable on this surface?

Appendix: Proof That ALG Is Well Formed and Finite

By reason of (1) and (3) the variable i can (re)enter (2) at most $2n + 3$ times. Hence both loops entering (2) are finite. Past (2) and (3) the value of i is in the range of $*$, so (4) is executable. Because $0 < 0^* = 2n + 1$, it follows from (3) and (4) that i enters (5) with $0 < i \leqslant 2n + 1$ and $i^* < i$. The initial value of the test pair (s, j) in the J-subroutine is $(-1, i)$. With some difficulty one sees that $j \leqslant i$ throughout the subroutine, until it either terminates at NO or returns to (A).

Moreover, $-1 \leqslant j$ at all times. For suppose, on the contrary, that instruction (6) advances j to something less than -1. This must happen in some cycling of the J-subroutine, and so we consider the first time this happens. Then, the test value at (B) read $(-1, -1)$. Tracing back, this quantity came through (11). Hence it left (6) previously as $(+1, -1)$ and therefore read $(+1, -2)$ at (B) contrary to the assumption. In fact, (s, j) can read $s = +1$ only if $j = -1$ at (B).

Thus $-1 \leqslant j \leqslant i$. Instruction (9) is feasible because past (8) $j \geqslant 0$ and therefore has a value j^*. Given (9), both (10) and (12) are executable. No cycling returns j to (B) with $j = i$. For suppose, on the contrary,, it does. Since (10) is the only place other than (6) where j changes value, it came through there. Then j entered (10) as $j^* = i^*$, contrary to the branching at (9).

We next show that every pair (s, j) at (B) coming from a cycling has a unique J-predecessor. If it has any predecessor, we have seen that $j \neq i$. If $j = -1$, it came through (11) because -1 has no * value. But then it left (6) as $(-s, -1)$ and the predecessor is $(-s, -1 + s)$. As has been shown in this case $s = +1$, and the unique predecessor is $(-1, 0)$. Next, we may assume that $-1 < j < i$, s arbitrary. Of the pair $\{j, j^*\}$, exactly one of the following can be said: both j and j^* are < 1; $j < i < j^*$; or $j^* = i$. In the first case, the pair must have branched south through (9) before it reached (B). Thus prior to (10) it read $(\nu(j^*)\, s.\ j^*)$ and the unique predecessor at (B) read $(\nu(j^*)\, s,\ j^* - \nu(j^*)\, s)$. In the second case, it must have branched west at (9), and the unique predecessor is $(s, j - s)$. In the third case, the unique predecessor is $(s, i^* - s)$.

Finally, it is seen that the J-subroutine is finite as follows. The pair (s, j) at (B) can read $(-1, i)$, $(+1, -1)$ or (s, j), s arbitrary but $-1 < j < i$, at most once. For suppose, on the contrary, there are repeated pairs. Let (s, j) be the first repeated one. It is not $(-1, i)$, as we have seen above. Hence both instances have a predecessor which, by assumption of priority, must differ. But this contradicts the uniqueness of predecessors.

References

1. G. K. Francis, The Folded Ribbon Theorem, to appear *Trans. Amer. Math. Soc.*, August (1969); summary in *Bull. Amer. Math. Soc.*, **74** (1968), 264–267.
2. C. F. Gauss, *Werke* **8**, 272, 282–286.
3. M. L. Marx, Normal Curves Arising from Light Open Mappings of the Annulus, *Trans. Amer. Math. Soc.* **120** (1965), 45–56.
4. J. v. Sz. Nagy, Über ein topologisches Problem von Gauss, *Math. Z.* **26** (1927), 579–592.
5. C. J. Titus, A Theory of Normal Curves and Some Applications, *Pacific J. Math.* **10** (1960), 1083–1096.
6. C. J. Titus, The Combinatorial Topology of Analytic Functions on the Boundary of a Disk, *Acta Math.* **106** (1961), 45–64.
7. L. B. Treybig, A Characterization of the Double Point Structure of the Projection of a Polygonal Knot in Regular Position, *Trans. Amer. Math. Soc.* **130** (1968), 223–247.
8. R. F. Verhey, Diffeomorphic Invariants of Immersed Circles, submitted to *Trans. Amer. Math. Soc.*; University of Michigan Dissertation, 1966.
9. H. Whitney, On Regular Closed Curves in the Plane, *Compositio Math.* **4** (1937), 276–284.